**Alberta Heritage Foundation
for Medical Research**

Intensive Intervention Programs for Children with Autism

Sue Ludwig, Christa Harstall

February 2001

Intensive Intervention Programs for Children with Autism

Sue Ludwig, Christa Harstall

February 2001

This Health Technology Assessment has been prepared on the basis of available information of which the Foundation is aware from public literature and expert opinion and attempts to be current to the date of publication. It has been externally reviewed. Additional information and comments relative to the report are welcome and should be sent to:

SUMMARY

- Intensive Behavioural Intervention (IBI) seems to be an umbrella-like term used to describe behavioural programs that are intensive (up to 40 hours per week) for autistic children. Included under the term IBI is Lovaas therapy.

- Three critical reviews of intensive intervention programs for autism by ECRI, BCOHTA and Smith were summarized. Of the three reviews, the one by ECRI was the most inclusive, analyzing studies on Lovaas therapy, TEACCH, the Rutgers Program, the Denver Program, LEAP and the Autism Pre-school Program.

- All critical reviews analyzed studies on Lovaas therapy and concluded that these studies were methodologically flawed.

- ECRI concluded that Lovaas therapy appears to increase scores on IQ tests and behavioural adaptation, at least in some children with autism; however given the studies' designs and methodological flaws, they could not determine if the changes in IQ and functional parameters could be attributed to the Lovaas therapy.

- BCOHTA concluded that the study conducted by Lovaas and the follow-up study done by McEachin and colleagues were methodologically stronger that other published studies; however, they were still inadequate to establish the degree to which this form of therapy resulted in "normal" children.

- Smith stated that methodological flaws in the research hinder the ability to draw conclusions; however, the studies by Lovaas and McEachin and colleagues had the strongest study design. Children with autism in these studies made "major, long-lasting improvements as a result of the treatment they underwent". Smith also stated that two of the three studies, which attempted to replicate the Lovaas study, produced favorable results.

- The outcome measurement instruments, used in all of the studies assessed in the critical reviews, were very similar. Most researchers employed standardized measures of IQ tests, adaptive functioning and language development.

- It appears that children improve in functioning with intensive intervention programs, but it remains to be determined if any one program is more effective than another.

- There is insufficient evidence to establish a relationship between amount (intensity and duration) of any intensive intervention treatment program and outcomes measures (intelligence tests, language development, adaptive behaviour tests).

ACKNOWLEDGEMENTS

CONTENTS

Tables:

INTRODUCTION

Alberta Children's Services, Alberta Health and Wellness and Alberta Learning requested the Health Technology Unit (HTA) of the Alberta Heritage Foundation for Medical Research (AHFMR) to review the published research on the effectiveness of intensive behavioural intervention (IBI) programs for children with an autism spectrum disorder (ASD). Recent critical reviews of intensive intervention programs for children with autism were conducted by other HTA agencies including Emergency Care Research Institution (ECRI) and the British Columbia Office for Health Technology Assessment (BCOHTA). A critical review was also conducted by Smith in 1999. In light of this recent work, it was decided that the HTA Unit of the AHMFR would summarize their analysis. In addition, a listing of the outcome measures from the primary studies included in these critical reviews would be detailed.

AUTISM

Autism is a developmental disability that usually presents before three years of age. It is a life long disability and etiology remains unclear (Ward, personal communication). Autism is a disorder marked by severe intellectual, social and emotional impairment. Children with autism may demonstrate poor response to sensory stimuli, not recognize their parents, and may lack interest in their environment [45]. According to the fourth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, the essential features of an autistic disorder are the presence of "markedly abnormal or impaired development in social interaction and communication, and a markedly restricted repertoire of activities and interests" [3]. These features must be present prior to age three (Smith, personal communication). The prognosis appears to be extremely poor [36].

DEFINING IBI

Intensive Behaviour Intervention appears to be a generic term referring to behavioural interventions that are intensive and comprehensive. Some have used the term IBI interchangeably with Applied Behavior Analysis (ABA) or with Lovaas therapy. The Autism Society [5] noted that the terms are not strictly synonymous because Lovaas states that only a practitioner trained and affiliated with his program can properly be said to deliver "Lovaas therapy". Lovaas did not refer to his treatment as either IBI or ABA; rather he called his program the "Lovaas UCLA Program".

The California Departments of Education and Developmental Services [12] defines IBI as "early intervention carried out all or most of the child's waking hours, addressing all significant behaviours in all of the child's environments by all significant persons for many years".

CRITICAL REVIEWS OF INTENSIVE INTERVENTIONS

Three critical reviews of intensive interventions programs were found in the literature (see Methodology, Appendix A). These were conducted by:

- ECRI in 2000 [15];
- BCOHTA in 2000 [6]; and
- Smith in 1999 [52].

ECRI [15] was the most inclusive in their review on comprehensive treatment programs for children with autism. The comprehensive treatment programs included were:

- Lovaas Therapy,
- The Rutgers Autism Program,
- The TEACCH Program,
- The Denver Model,
- The LEAP Program, and
- The Autism Pre-school Program.

BCOHTA [6] published a review of Lovaas therapy for children with autism. Their critical review focused primarily on Lovaas therapy; however, they also critically reviewed one study conducted on the effectiveness of the TEACCH Program [37].

Smith [52] critically reviewed nine studies on behavioural therapy, one on the TEACCH Program and two on the Denver Model.

ECRI Report [15]

ECRI focused their technology assessment on six comprehensive programs summarized in Table 1.

Table 1:   Programs examined by ECRI

Program	Description	Intensity	Cost* American Dollars
Lovaas therapy	Behavioural	40 h/week	UCLA - $810/day Non-Local - $1,350/day May Institute - $100,000/year (residential) $30,815/year (young child program) $40,000/year (home-based)
Rutgers	Behavioural	40 h/week	$1,000/day (6 hours/day)
TEACCH	Developmental	30 h/week	State Funded
Denver	Developmental	22-35 h/week	Not available
LEAP	Developmental/ Behavioural	15 h/week	$25,000/school year (1991-1992)
Autism Pre-school Program	Developmental/ Behavioural	Unspecified	Not available

One of the aspects that all of the above programs have in common is their focus on early intervention.

A brief summary of other treatment options for children with ASD is provided in Appendix B.

A.   Lovaas Therapy

Description

Lovaas therapy, developed by O. I. Lovaas, utilizes time-intensive **behavioural** intervention techniques. The program is intended to treat children with autism between the ages of 2 to 3 years. Treatment is provided initially in the home. Lovaas therapy uses principles of operant conditioning - positive behaviours are reinforced and negative or aggressive behaviours are ignored (extinguished). More acceptable forms of behaviour are taught. Originally Lovaas therapy used punishments such as electric shock; however, this is no longer practised. The program gradually teaches self-help and receptive language skills, nonverbal and verbal imitation skills and establishes the foundations of appropriate play.

An essential aspect of Lovaas therapy is teaching imitative skills to the child. Once learned, imitation is used as one of the prime teaching tools. The second stage of the program begins once the child has mastered basic skills and involves teaching of expressive and early abstract language and interactive play. The child is taught at home (and at school in later stages) and advances on to learn early academic tasks, socialization skills, cause and effect relationships. Each child is taught on a one-to-one basis for up to 7 hours per day (40 hours/week). Play sessions and breaks are dispersed throughout the day.

For the first 6 to 12 months, one-to-one home based teaching is provided. Teaching gradually becomes less structured as the child is prepared to attend school. Student aides accompany the child at school to facilitate the transition from home-based to classroom techniques. The student aide is gradually phased out.

Studies of effectiveness

ECRI critically appraised five studies on the effectiveness of Lovaas therapy. The McEachin and colleagues [36] study was a follow-up to the original Lovaas study [35]. Table 2 summarizes the studies ECRI analyzed along with their critique of each of the primary studies.

Table 2: Studies of Lovaas Therapy

Study	Sample, and Outcome of Study	Critique by ECRI
Lovaas [35] 1987 prospective/controlled Pseudorandomized. McEachin, Smith and Lovaas [36] 1993 follow-up to the Lovaas 1987 study.	Sample: N = 19 in TG N = 19 (+21)* in CG Outcome: TG – 9/19 = normal functioning CG – 0/19 = normal functioning	▪ Sample assigned to groups based on staff availability; pseudorandomized ▪ Small sample size, low statistical power ▪ Different instruments used to measure IQ pre & post, but the same instruments were used for both groups ▪ Groups were matched across 20 different domains ▪ Drop-out bias weakened validity ▪ Sample not representative of autism population (children on medications not included, ratio of males to females not representative of that in population).
Anderson et al. [4] 1987, single group, pre-post design	Sample: N = 14 Outcome: 2 year follow-up – children average increase of 20-22 IQ points over intake on standardized tests.	▪ Different instruments used pre & post except the Uniform Performance Assessment System that was used both times. ▪ Pre-post study design is vulnerable to biases that can affect the validity of conclusions drawn between treatment and outcome. ▪ Maturation bias possible. ▪ Children on medications were not excluded. ▪ Drop-out bias weakened validity. ▪ Only 6 children participated in the follow-up. ▪ Despite attempting to replicate the proscribed Lovaas program, only 15-25 hours/week of treatment were provided.
Birnbrauer and Leach [9] 1993 nonrandomized, controlled	Sample: N = 11 TG N = 8 CG Outcome: TG – no normal functioning; 4/9 achieved IQ>89; the remaining declined.	▪ Did not report if children were on medication. ▪ Drop-out bias weakened validity (38% of CG and 18% of TG). ▪ No data for follow-up were presented. ▪ Despite attempting to replicate the proscribed Lovaas program, only 19 hours/week of treatment were provided.

Table 2: Studies of Lovaas Therapy (cont'd)

Study	Sample, and Outcome of Study	Critique by ECRI
Sheinkopf and Siegel [50] 1998, case-control retrospective study	Sample: N = 11 TG N = 11 CG Outcome: TG – no normal functioning; mean IQ increase of 25 points.	▪ Did not report if children were on medications. ▪ Hours of special education was not equal (CG=10.71, TG=6.41). ▪ Groups matched on IQ, age, # of symptoms, but not on other variables such as language ability. ▪ Some children received additional professional therapies (OT, speech therapy) and the amount differed between groups (TG=1.16 hours/week CG=.44 hours/week). ▪ No data for follow-up were presented. ▪ Drop-out bias weakened validity. ▪ Despite attempting to replicate the proscribed Lovaas program, only 20 hours/week of treatment were provided.

TG = treatment group
CG = control group
* an additional control group composed of 21 children from another study.

Conclusions

Based on the quality of the studies reviewed by ECRI, the following question can be answered: "Is some improvement occurring?", but not the question, "Can any or all of this improvement be attributed to Lovaas treatment?" The Lovaas [35] study results indicated significant improvement by the treatment group versus the two control groups in both IQ and school placement. ECRI stated that "although the original study of Lovaas cannot be considered reliable", the follow-up study by McEachin et al. [36] in 1993 indicated that improvements associated with Lovaas therapy were maintained by the treatment group for an average of 5 years in terms of IQ and school placement.

The significant improvement in IQ reported by McEachin et al. [36] was further supported by results obtained from the study conducted by Sheinkopf and Siegel [50]. ECRI stated that "although this study was not well controlled for, the lack of difference of the groups in terms of IQ was confirmed and therefore the results on this measure may be reliable".

ECRI's analysis indicated that children who received treatment based on Lovaas therapy showed improvement in IQ, even when the treatment program was applied less intensively than Lovaas originally recommended. ECRI also concluded that the "studies provide evidence that improvement did occur on certain functional parameters, including a mean reduction in maladaptive behaviours and symptom severity and an increase in socialization and daily living skills". However, these improvements were based on parents report of changes and the parents were not blind to the treatment conditions (Gresham, personal communication).

Given the methodological issues identified by ECRI, they could not "definitively determine that Lovaas therapy per se caused the changes in IQ and functional parameters". The available evidence is only suggestive of treatment effectiveness.

B.   The Rutgers Autism Program

Description

This program is an early intensive **behavioural** intervention program delivered by doctoral psychologists and behaviour specialists/analysts. It is a home-based program provided on a full-time basis for at least 2 years. In many aspects it is similar to Lovaas therapy. It differs in that it does not provide the staff to deliver the treatment. Instead, program staff consult with families and schools and provide training in the implementation of the program and for follow-up. The families either provide their own treatment or hire staff trained by the program. The Rutgers program assists in preparing the child for placement into the classroom while gradually withdrawing home-based treatment.

Studies on effectiveness

ECRI presented one study by Weiss in 1999 that evaluated the effectiveness of the Rutgers Autism Program [57]. This study was a single group pre-post design with a sample of 20 children. ECRI concluded the results of the study could not be used to answer any questions about effectiveness.

C.   The TEACCH Program

Description

This is a statewide program based in Northern Carolina for autistic children. It is a **developmental** program established on the assumption that certain developmental skills are prerequisites to learning. The objectives of TEACCH are to:

- "maximize adaptation through structured teaching of new adaptive skills;
- to develop environmental modifications to accommodate the child's deficits;
- to maintain close collaboration between teacher and parent;
- to provide a continuity of structured teaching throughout life; and
- to prevent the development of further behavioural problems" [6].

TEACCH uses structured teaching as its principle technique. The classroom is structured physically in scheduling and in teaching methods to provide continuity for the child. The children with autism are in classes with other children with developmental disabilities. Each child has his or her own workstation and clear indications of where each educational activity will occur. For example, writing is done at tables; toilet training and eating are done in special self-care areas.

Studies on effectiveness

ECRI presented three studies about the effectiveness of the TEACCH program [37, 38, 51]. These are summarized in Table 3.

Table 3:   Studies of the effectiveness of TEACCH

Study	Sample and Outcome of Study	Critique by ECRI
Ozonoff and Cathcart [37] 1998, pseudorandomized, controlled trial	Sample: N = 11 TG N = 11 CG Outcome: TG – improved significantly more than the CG on scores on the PEP-R	▪ Small sample size ▪ Potential bias due to extraneous events (such as the day program the children attended). ▪ Did not describe groups in sufficient detail ▪ Group assignment not randomized and not matched. ▪ Study originally not designed to assess the effectiveness of TEACCH, but to assess the effectiveness of a TEACCH-based home program as a supplement to the child's normal day treatment program (which differed in intensity between groups).
Short [51] 1984 Single group, AB design (a design in which children acted as their own control. All children had outcome measures taken at the end of an untreated control phase and then after the treatment phase).	Sample: N=20 Outcome: Treatment effect measured for all four main variables (parental guidance, appropriate child behaviour, inappropriate child behaviour & stress) were significant	▪ Maturation bias possible ▪ Did not control for extraneous events (such as children on medications were not excluded). ▪ Did not describe sample well ▪ Used non-validated measures (threatening construct validity). ▪ The effect of drop-outs were not analyzed.
Panerai, Ferrante and Caputo [38] 1997, Single group, pre-post design	Sample: N=18 Outcome: Not mentioned by ECRI	▪ Maturation bias possible ▪ Did not control for extraneous variables (did not exclude children on medications). ▪ Did not provide information on diagnostic criteria. ▪ Used non-validated measures ▪ No details of analysis provided, other than p values. ▪ Generalization to wider population unclear.

TG = treatment group
CG = control group
PEP-R =Psychoeducational Profile - Revised

Conclusions

ECRI concluded that the Paneria and colleagues [38] study could not be used to answer questions on effectiveness due to a lack of methodological details. The study by Short [51] had a high drop out rate (25%), causing ECRI to question the reliability of the results. They concluded that it could not be used to answer questions of effectiveness. Finally, the Ozonoff and Cathcart [37] study was

pseudorandomized and ECRI found it reasonable to assume that "the effects of biases such as maturation, test practice and regression bias were minimized in this study". ECRI concluded that this study could be said to identify whether a behavioural change had occurred, however, they could not pinpoint the cause of the behavioural change to the treatment.

D. The Denver Model

Description

This is a **developmentally** based program provided by the University of Colorado Health Science Center in Denver. The Denver Model is a joint educational and therapeutic program for autistic children aged 2 to 6 years. In 1993, the program was invited to join a public school system in its efforts to provide inclusive pre-school education for children with autism, and now it also provides comprehensive home-based programs.

Like TEACCH, the Denver Model aims to develop the capacities of the child with autism. Treatment is individualized and uses feedback from parents and the treatment team (consisting primarily of the parents and multidisciplinary members). The treatment plan includes goals and objectives, instructional plans and activities, and data collection (both quantitative and qualitative), and is used in all settings.

The treatment plan focuses on the development of communication and play skills, sensory activities, personal independence, and reducing unwanted behaviours. The attempt to develop communication skills is a dominant aspect of the program. This is accomplished through teaching, elicitation and shaping. The Denver Model also emphasizes the development of play skills as a part of normal development. The child's sensory systems are regarded as important and the program involves the child in "sensory-social activities" along with structured teaching activities. Child independence is also valued and independent goal-oriented tasks that contribute to the family are targeted. Finally, to minimize unwanted behaviour the program uses functional behavioural analysis, communication training, positive teaching of alternatives, and redirection to provide new acceptable behaviours.

Studies of effectiveness

ECRI reported that support for the effectiveness of the Denver Model comes from three studies [40, 42, 43]. These studies are summarized in Table 4. None of the studies used a control group and none satisfactorily controlled for confounders such as maturation bias or drop-outs. Therefore, ECRI concluded that these results could not be used to answer questions of effectiveness.

Table 4: Studies on the effectiveness of the Denver Model

Study	Sample and Outcome of Study	Critique by ECRI
Rogers and DiLalla [40] 1991, single group, pre-post-test	Sample: N=49 Outcome: Significant effects were demonstrated in 5 of 6 areas : perceptual/fine motor, cognition, language, social/emotional, and gross motor.	▪ Did not exclude children on medications. ▪ Pre-post-test technique used (prediction index analysis) has not been validated with children with autism and therefore results could be over or under estimates. ▪ Did not account for missing data.
Rogers, Lewis and Reis [43] 1989, single group, pre-post-test	Sample: N=31 Outcome: Significant treatment effects in cognition, perceptual/fine motor, social/emotional, and language skills were demonstrated.	▪ Did not exclude children on medications. ▪ Did not account for missing data on 9 children.
Rogers and Lewis [42] 1987, single group, pre-post-test	Sample: N=11 Outcome: Not mentioned by ECRI.	▪ Did not exclude children on medications.

E. The LEAP Program

Description

This is a comprehensive pre-school service system designed to meet the needs of both normal and autistic children. It was developed by Philip Strain and colleagues in Pittsburgh and is **developmentally** based with some **behavioural** analysis incorporated into it. The program provides training for parents and they are involved with the program, which is administered 5 days per week, 3 hours per day in a classroom setting. One-to-one intervention is not provided. Instead, teachers (with master's degrees) and an assistant provide intervention to ten normally developing children and three to four autistic children at a time. In addition to the teacher, a full-time speech therapist, occupational therapists, and physical therapists work with the children.

Individualized curriculum plans are developed with input from parents. These plans include short-term objectives and are updated every three to four months. The classroom is carefully arranged into theme areas to help the children learn. The schedule is designed to provide a variety of activities. The curriculum is supplemented to assist autistic children to learn language and functional skills, independent play and work skills, social interaction skills, and adaptive behaviour. Positive behavioural intervention procedures are used to help children who display aberrant behaviours.

School instruction is based on the "Tri-I (Innovative; Integrative; Individualized) curriculum for mainstreaming (TRIIC)". This program involves individualized learning in an integrated environment. It deviates from traditional one-to-one

instruction for children with autism. ECRI noted that no evidence exists to support the hypothesis that one-to-one instruction is superior to group instruction or visa-versa, even though Lovaas argued that any treatment protocol that does not use individualized instruction is inappropriate for autistic children.

Studies of effectiveness

Only one study [27] on the LEAP program was reported by ECRI. This study was a single group, pre-post design of 13 children with autism. ECRI concluded that the small sample size, the confounding effects of maturation and the poor generalizability of the sample precludes the use of this study to answer questions of effectiveness.

F. The Autism Pre-school Program

Description

This is a collaborative program based at the University of Manitoba. It is staffed by a multidisciplinary team and requires collaboration between the university hospital, provincial government, and local community resources. It uses a variety of standard **behavioural** and **language development** methods and is similar to the Rutgers Autism Program in that it operates primarily in a consultative role. Intervention occurs through parents and day-care staff and some of the treatment is performed in an integrated day-care setting.

Caregivers are taught to understand and empathize with the child, perform a functional behavioural analysis and plan, and evaluate strategies for changing behaviour. The development of language and social skills is given priority over individual behavioural problems.

Studies of effectiveness

ECRI reported one study [28] that examined the effectiveness of the Autism Pre-school Program. This was a randomized, controlled trial involving a sample size of 36 children with autism. The study design minimized threats to the validity of the results. It demonstrated strong construct validity and ECRI concluded that the results are generalizable to pre-school children with autism. The treatment group in this study demonstrated a statistically significant improvement in language development over the control group. No other outcome measures were significant. The treatment period of this study was short, however, only over 12 weeks. This period of time was not likely long enough to assess the effects of the program on overall symptoms of autism.

Summary of ECRI's Analysis

In total, ECRI analyzed 18 studies that assessed the efficacy of comprehensive treatment programs for autistic children. Four of the original 18 studies were excluded from further analysis, leaving 14 studies. Of the 14 studies, only six contained a control group and only one of these studies was a randomized controlled trial. The other eight studies were single group, pre-post study designs.

ECRI noted that all studies assessed reported positive treatment outcomes. However, they also stated that all of the studies were methodologically weak, and this "diminished to greater or lesser degrees the value of the reported data".

ECRI concluded that the children who participated in the studies that evaluated the effects of Lovaas therapy, appeared to improve. However, due to methodological limitations, it is not certain whether, or to what extent, this improvement could be attributed to the Lovaas therapy. They also stated that "it is difficult to ascribe these improvements to anything else".

For the Autism Pre-school Program, ECRI concluded that the children appeared to improve in the study, and that because of study design it was suggested that the improvement was due to the program. However, they noted that while the children improved in language development, they did not show improvement in other measures such as cognition, motor skills or IQ. They stated that the lack of improvement in these areas may be due to the study sample being too small and the study follow up too short.

In regards to the TEACCH program, ECRI concluded that although statistically significant improvements in imitation, perception, fine motor skills, gross motor skills and cognitive performance were found, the functioning of the control groups was not that different from the experimental group. Therefore, they noted the meaning of the results in practical terms is uncertain.

Overall, ECRI concluded:

> "It does appear possible to improve some aspects of function in autistic children, but it is not clear that any one program is more effective than another. The possibility that any positive, intensive efforts directed toward the child will bring about improvement, regardless of whether those efforts are associated with any specific program or any formal program at all."

BRITISH COLUMBIA OFFICE OF HEALTH TECHNOLOGY ASSESSMENT REPORT [6]

BCOHTA published a critical appraisal of the effectiveness of Lovaas therapy in June 2000. The research question they sought to answer was "what is the effectiveness evidence that early, intensive behavioural treatment programs for pre-school children with autism results in improved overall outcome versus alternative management strategies?" BCOHTA reviewed the same studies as did ECRI with the exception of the Anderson and Glynnis [4] study. They also reviewed the same study [37] on the TEACCH Program as did ECRI. Table 5 summarizes their analysis and critiques of the studies further to those already provided by ECRI.

Table 5: Lovaas Therapy and critiques by BCOHTA

Study	Additional Critiques by BCOHTA
Birnbauer and Leach [9] 1993	▪ The study was too small, too short and methodologically weak.
Lovaas [35] 1987 McEachin, Smith and Lovaas [36] 1993	▪ Lovaas studies did not compare two different therapies, they only compared two levels of intensity of behavioural intervention. There were insufficient details to determine what therapies were provided for the second control group. ▪ Group assignment was not random. Sampling was not random; therefore generalizability cannot be assured. ▪ Children in the sample were higher functioning than most autistic children and not representative of sex ratio in the autistic population. ▪ The extent to which outside evaluators were used is questionable – school performance and behavior was rated by parents & teachers. Outside evaluation was limited to IQ testing of 9 children from the active treatment group. ▪ Improvement in IQ might reflect improvement in compliance with test taking rather than in cognitive functioning. ▪ There was a lack of details in actual adherence to treatment protocols and for alternative therapies the experimental group received (e.g.: medications). ▪ School placement was used as an outcome measure. This may reflect the policies of the school system towards special needs children rather than changes in the child. ▪ Many other skills necessary for normal functioning, such as social interaction were not measured. ▪ No details were provided about when the children started treatment, how long the treatment was provided, and when follow-up assessment occurred. ▪ The validity of the scaling procedure for prorated mental age is questionable as details to determine the individual child's age, mental age and how they were combined with IQ tests were not provided.

Table 5: Lovaas Therapy and critiques by BCOHTA (cont'd)

Study	Additional Critiques by BCOHTA
Ozonoff & Cathcart [37] 1998	• The intervention period of 10-12 weeks was too short. • The study was not methodologically sound.
Sheinkopf and Siegel [50] 1998	• Retrospective, poorly designed study that was too unreliable to draw even weak support in terms of IQ benefits.

Conclusions by BCOHTA

BCOHTA concluded:

"while many forms of behavioural therapy clearly benefit children with autism, there is insufficient valid effectiveness evidence to establish a causal relationship between a particular program of intensive, behavioural treatment, and the achievement of normal functioning".

Smith, T. – Outcome of Early Intervention for Children with Autism [52]

Smith conducted the third critical appraisal of nine studies on early interventions (Lovaas therapy [4, 9, 35, 36, 50], the Rutgers Program [21, 23, 22], the LEAP Program [27], and the Princeton Child Development Program (PCDI) [16]. Under "other programs" he also reviewed two studies on the Denver Model [40, 41] and one study on TEACCH [33] (not the same study as either ECRI or BCOHTA). All of these programs (Table 1) have already been presented in previous sections except for the PCDI program. Smith described the PCDI program as a program that offers 27.5 hours per week of **behavioural** intervention in multiple settings.

Smith noted that most reports of major gains made by children with autism have "withered under scrutiny" as concluded in the study by Handleman and colleagues [21]. He added few additional critiques of the Lovaas studies from those already mentioned by ECRI and BCOHTA.

He did add that most studies (except the McEachin et al. [36] follow-up study of the original Lovaas study) did not provide data on the children's progress following termination of treatment. Smith noted that this is a critical omission because even if treatment is successful while ongoing, progress may not continue when the children are not receiving specialized services. The need to validate long-term benefits is important in outcome studies.

Smith also added that in the studies he reviewed only two [21, 23] relied on a standardized diagnostic instrument, the Childhood Autism Rating Scale (CARS). In both of these studies some children did not meet the diagnostic criteria for an ASD, even though the CARS produces more false negatives than most other measures. Thus, he noted, in these, and perhaps other, studies, some children may not have met generally accepted criteria for an ASD.

Smith concluded that methodological weaknesses in the research hinder us from drawing conclusions from existing early intervention studies. Of those he reviewed, the UCLA studies [35, 36] examining the effectiveness of Lovaas therapy had the most "favorable results and the strongest methodology". He noted that even though these studies created much controversy, there is general agreement that as a result of treatment the children made major, long-lasting improvements. Two partial replications of the original UCLA study have also reported favorable results [4, 50]. Another replication study [9]; however, reported mixed results. Smith stated that:

> "the extent to which most exact replications would obtain results comparable to those of Lovaas [35] remains an open question of considerable importance".

Summary of the Three Critical Appraisals

ECRI chose 14 research studies for their analysis in order to determine the efficacy of comprehensive treatment programs. BCOHTA analyzed four research studies on early intensive behavioural therapy (Lovaas therapy) and the same study as ECRI on the TEACCH Program. Different primary research studies were chosen for inclusion in the critical reviews by BCOHTA and ECRI. This is due to the different methodology used for the critical appraisals. ECRI included all studies on comprehensive programs for autism therapy and looked for associated evidence. BCOHTA included only studies with control groups since these research designs were deemed to be appropriate to draw efficacy conclusions (Bassett, personal communication).

Smith analyzed nine outcome studies of behavior programs (including the Lovaas studies), one study on the TEACCH Program (although this was not the same study as the one included in the other two critical reviews), and two studies on the Denver Model.

ECRI stated that all of the studies reported positive treatment outcomes. At the same time, all suffered from methodological weaknesses that diminished the validity of their conclusions. They wrote:

> "Choosing a particular program for a child with autism presents parents with no small amount of difficulty. The results of this technology assessment indicate that there are interventions that are associated with some improvement in autistic symptoms. However, claims of turning autistic children into children indistinguishable from normal children have not been substantiated. …. Currently there does not appear to be a "magical" intervention."

BCOHTA concluded that the Lovaas [35] study and follow-up study by McEachin and colleagues [36] were methodologically sounder studies than published reports of alternative comprehensive therapies (TEACCH); however, they also stated that:

> "there is insufficient effectiveness evidence to establish a relationship between the amount (per day and total duration) of any form of early comprehensive treatment program and overall outcome"

Finally, Smith concluded that methodological weaknesses in the studies he reviewed hinder us from drawing firm conclusions from existing research. He agreed with both ECRI and BCOHTA that the Lovaas studies [35] had the most favorable results and the strongest methodology. Even so, he noted the evidence is preliminary and the claims made by the researchers still need to be replicated by independent researchers using improved methodologies.

OUTCOME MEASURES

In the studies reviewed by ECRI, BCOHTA and Smith, most focused primarily on measures of intellectual functioning to determine outcomes. Some studies included parent ratings and school placement as additional outcomes. The list of outcome measures from the individual primary studies are provided in Table 6 (Appendix C). Appendix D provides a more comprehensive list of possible measures and tests useful for children with ASD.

A description for each outcome measure listed in Table 6 was not available. The following provides the available information about each specific measurement instruments as described by the researchers and supplemented with one of the most widely used references on the assessment of children [45].

Intelligence tests

Many of the studies used variance of intelligence as an outcome measure. Tests to measure intelligence in the studies included the Bayley Scales of Infant Development, Stanford Binet Intelligence Scales, the Stanford-Binet IV, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), the Wechsler Intelligence Scales for Children-Revised (WISC-R), Merrill-Palmer Scale of Mental Tests (MPSMT), and the Leiter International Performance Scale (LIPS). Sattler [45] noted that the most common choices for testing intelligence of children with autism include the LIPS and performance scales of the WISC-R, and Stanford Binet. All of these intelligence tests are standardized, valid and reliable instruments, and are commonly used in research studies as well as for clinical purposes.

Criticisms have been levied against Lovaas for the instruments chosen to measure outcomes in the UCLA Project, as well as procedures used to administer the assessments. Gresham and MacMillan [20] stated that "any changes in measuring instruments, administration and scoring procedures, test administrators, or scaling can affect the results of an experiment". In the Lovaas study the children were pre-tested with a wide variety of measurement instruments such as the Stanford Binet, the Bayley Scale of Infant Development and Cattel Infant Intelligence Scale. These children were then tested again 3 years later with the WISC-R or the LIPS.

Gresham and MacMillan [20] stated: "the children were post-tested with different measures than they received at pretest, thereby making the results of these different testings uninterpretable". One needs to examine the correlations between the tests used at pretest and posttest. Sattler [45] reported that the correlations between the WISC-R and the Stanford Binet IV to be around 0.78 and stated that the WISC-R and the Stanford Binet likely yield composite scores similar to each other. The Stanford Binet (norms after 1972) correlates with the

WPPSI at around 0.80. However, Sattler also noted that the above conclusions may not hold for special populations such as autistic children.

Adaptive behaviour and behaviour problems

The most common measure of adaptive behaviour and behaviour problems in the studies was the Vineland Adaptive Behavior Scales (VABS). This scale assesses the social competence of handicapped and non-handicapped persons from ages 0 to 19 years. Adaptive behaviour is measured in four domains, including communication, daily living skills, socialization, and motor skills. In each domain a standard score is calculated. The VABS is standardized. Reliability ranges from 0.83 to 0.90. The VABS correlates at 0.52 with the WISC-R. Sattler [45] stated that it is a potentially useful tool for the assessment of adaptive behaviour but difficulties still exist with the scales. He also noted that the "norming procedures created serious fluctuations in the means and standard deviations of the standard scores from age group to age group", and is a particular problem with mentally retarded children.

Tests of language performance and development

In the studies reviewed various tests of language ability and development were used for outcome measures, including the Preschool Language Scale, Battelle Developmental Inventory (BDI), the Learning Accomplishment Profile (LAP), the Reynell Developmental Language Scales, the Uniform Performance Assessment System, the Peabody Picture Vocabulary Test, and the Early Intervention Developmental Profile and Preschool Profile.

The BDI is an assessment battery of five domains of development: personal-social, adaptive, motor, communication, and cognitive in children up to 8 years of age. Reliability and validity of the BDI is high. It has a high correlation with the VABS, the WISC-R and the Stanford Binet. Sattler [45] stated that it is a useful measure to obtain information about important areas of development in young children. The usefulness of the BDI for assessing autistic children was not discussed.

The Reynell Developmental Language Scales provides a measure of comprehension and language. It is a common assessment instrument used in research studies [39]. No information about the reliability/validity was provided in the study that used it [9], or by Sattler [45].

The Psychoeducational Profile – Revised (PEP-R) [37] was developed by professionals of the TEACCH Program. It is an inventory of behaviours and skills designed to identify uneven and idiosyncratic learning patterns of children with autism. It provides information on developmental functioning in imitation, perception, fine motor, gross motor eye-hand integration, cognitive performance, and cognitive verbal areas. Information about the reliability and validity of this instrument was not reported by Schopler and colleagues [46].

The Early Intervention Developmental Profile and Preschool Profile assesses development in areas of language, cognition, fine motor/perceptual, gross motor, social/emotional, and self-care [28, 40, 41, 42, 43]. Rogers et al. [43] noted that studies have been carried out on this instrument with both handicapped and non-handicapped individuals. Concurrent validity of the profile scales has been shown to be high (correlations between 0.62 and 0.97) with other developmental measures such as the Bayley Mental and Motor Scales, and the Vineland Social Maturity Scale. Test-retest reliability figures have demonstrated scale score stability (reliability co-efficients ranged from 0.85 to 0.95).

The Uniform Performance Assessment System is a norm-referenced assessment instrument that lists skills (communication, social/self-help, pre-academic, motor and behavioural) in a normal developmental sequence [4]. No information was provided by the authors [4] about reliability/validity of the measure.

The LAP is a measure that yields developmental levels in eight domains; fine motor writing, fine motor manipulation, language naming, language comprehension, cognitive matching, cognitive counting, gross motor object movement, and gross motor body movement. No information about psychometric properties were given by the researchers who used it in their study [27].

Personality assessment

In the studies reviewed only two [9, 36] included an outcome measure of personality. The instrument used was the Personality Inventory for Children. Sattler [45] did not provide information about this test.

Autism rating/assessment scales

Four studies used the Childhood Autism Rating Scale (CARS) [28, 37, 40, 43]. The CARS is a rating system containing 15 scales that rate the severity of various aspects of autism on a continuum of one to four. The scales include relationships with people, imitation, affect, use of body, relation to non-human objects, adaptation to environmental change, visual responsiveness, auditory responsiveness, near receptor responsiveness, anxiety, verbal communication, nonverbal communication, activity level, intellectual functioning, and general impression. It is also a measure that was developed by the TEACCH program staff. No information about the reliability/validity is provided in any of the studies or by Sattler [45].

Jocelyn et al. [28] also used the Autism Behavior Checklist (a 57-item questionnaire to measure a rater's perception of a child's functioning on five subscales) and the TRE-ADD Autism Quiz (true/false questionnaire to measure parent's knowledge), however psychometric properties of either instrument were not available.

Parent measures

Only one study used the Parenting Stress Index [9]. The researchers did not describe it in their study; however it is a standardized, widely used test with good reliability and validity. It is administered to parents to identify stressful areas in parent-child interactions.

COMMENTS ON OUTCOME MEASURES

Sattler [45] noted that the handicaps of autistic children, such as their difficulties in establishing social relationships, their impaired communication skills, and their unusual responses to sensory stimuli, create difficulties in testing them. A thorough assessment requires an evaluation of language severity, intensity and frequency of maladaptive behaviours, intelligence, neurological status and familial factors. He cautioned that many children with autism may lack the sustained attention and linguistic skill that many intelligence tests require and deficiencies in test scores may reflect this rather than true abilities.

The Lovaas [35] and McEachin, Smith and Lovaas [36] studies were the best research design, randomized controlled clinical trial. However, the children were pre-tested with five different measures of intelligence/development and then the same children were post-tested with different measures such as the WISC-R 3 years later. Many of the pre-test instruments (Bayley Scale of Infant Development, Cattell Infant Intelligence Scale) have an extremely low correlation (0.10 to 0.20) [45] with the WISC-R (Gresham, personal communication). Measuring study effects in this manner results in threats to the internal validity of the study.

Another argument which advances the need to use the same instruments pre and post test is statistical regression effects (Gresham, personal communication). Statistical regression effects should be based on the stability coefficient between the pre and post test of the same instrument, not on the correlation between different measures taken at the same time (Gresham, personal communication). Gresham and MacMillan [20] provide a good overview of the impact that statistical regression may have had on the Lovaas [35] and McEachin et al. [36] studies.

A major threat to internal validity is instrumentation differences within the instrument itself as it evolves (Gresham, personal communication). This refers to changes over time in measurement instruments. For example, persons taking newer editions of IQ tests score differently on these than on previous older editions of IQ tests due such changes as variances in standardization (changes in the measured intelligence of the population over time).

Most of the primary studies conducted indicated that most comprehensive treatment programs showed similar gains with respect to IQ. These positive outcomes in relation to IQ across the across different treatment models (except for the Autism Pre-school program) may be a result of common program elements such as parent involvement/training, predictable routines and functional approach to problem behaviours (Gresham, Ward, personal communications). Furthermore, using school placement along with IQ tests as an indicator of effectiveness, is questionable. Many other factors influence the

determination of whether students with ASD enter regular or special classes (Calder, personal communication).

The key issue regarding outcomes measures may not be the measures themselves but the importance of not focusing on only one or two measures. Multiple independent measures should be administered by the same assessors (blinded to group assignment) at the beginning of the study and at the end of the study to all children enrolled in the study (Bassett, personal communication).

Conclusion

The three critical reviews evaluated a number of comprehensive treatment programs for young children with ASD. These included programs ranging from strict operant discrimination learning (Lovaas therapy) to broader applied behaviour analysis such as the Rutgers Autism Program to more developmentally oriented programs such as the Denver Model and TEACCH Program. Furthermore, these treatment programs vary in their intensity from 40 hours per week for Lovaas Therapy and Rutgers Autism Program to a range of 15 hours per week for the LEAP Program. Most of these intensive interventions were shown to be effective in producing developmental gains, increases in IQ and less restrictive school placement.

Because of the methodological limitations and weaknesses of existing research, evidence remains limited on the efficacy and effectiveness of one intervention in comparison to another. It does appear that children improve in functioning (as measured by various indices) with behavioural intervention programs. However, it remains to be determined if any one program is more effective than another program.

Generally the researchers used standardized, reliable and valid measures of intellectual functioning, adaptive behavior, language performance and development, and various measures of personality, autism rating scales and stress measures. The most common outcome measure used in the studies was that of IQ. Sattler [45] stated that some autistic children do not differ intellectually from other "normal" children, while some show unevenness in functioning. According to the DSM-IV, these children demonstrate impaired social interaction and communication, as well as diminished interest in their surroundings. Therefore it seems that to truly demonstrate efficacy, outcome measures would need to demonstrate improvements in these areas, rather than rely (for the most part) on IQ measures.

Further well designed research studies using multiple independent measures are required. These measures should be used for all children enrolled in the study, administered by the same assessors blinded to the intervention, and then used again for all children at the end of the study. Optimal intensity and duration of treatment for children with ASD remains to be determined through well designed studies.

More studies on the impact of these interventions on family members may be useful. Studies that look at changes in attitudes of family members may increase the family's ability to live with an autistic child. These types of studies may also identify 'system changes' that would enhance quality of life for both the family members and the child with ASD.

APPENDIX A: METHODOLOGY

HTA Search Strategy & Results

Search strategy for reviews or assessments of autism treatments and programs

The following databases were searched for reports/assessments related to autism, using only the search word 'autism':

- CCOHTA
- BCOHTA
- Cochrane Database of Systematic Reviews
- NHS Centre for Reviews and Dissemination (includes three databases – Health Technology Assessment Database, NHS Economic Evaluation Database, and Database of Abstracts of Reviews of Effectiveness.
- INAHTA
- ISTAHC
- NCCHTA
- US Congress – Office of HTA (archived site)
- NZHTA New Zealand Health Technology Assessment: The Clearing House for Health Outcomes and Technology Assessment
- AHRQ (formerly AHCPR)
- HSURC
- MCHPE
- ICES
- NIMH

In searching the NHS Centre for Reviews and Dissemination databases, **one** review done in France (it is in French, and it does not appear that the abstract is even in English). Another one was completed in the Netherlands, and its focus was on behavioural treatments for individuals with "mental retardation", and they included autism in their participant inclusion criteria. The **NIMH** was searched and some related documents were found, but not reviews or assessments. All the other databases returned no documents matching 'autism'.

A search of the other databases was undertaken as outlined below.

Database Searched *	Subject Headings	Textwords	Results
MEDLINE 1985- May 2000	Exp autistic disorder	educat$ OR treat$ OR intervention$ OR therap$ OR service$ AND Lovaas OR teacch OR denver OR leap OR rutgers OR pre-school OR pre-school	35
PreMEDLINE (As of July 31, 2000)		(autism OR autistic) AND (service* OR intervention* OR treat* OR educat* Or program* OR therap*)	1
EMBASE 1990-1999	Exp autism Exp infantile autism	Same as MEDLINE search	21
Best Evidence 2000 1st quarter		Autism	0
HTA		Autism	3
EED			
DARE			
Cochrane Database of Systematic Reviews			
ISTAHC database		Autism	1
HealthSTAR (1985-April2000)		Autism	6
PsycInfo (1985- April 2000)	Exp autism	Same as MEDLINE search	24
CINAHL	Exp autism	Same as Medline search	31
ERIC 1985-March 2000)		Same as Medline search	6
Dissertation Abstracts		Autism and (treatment or rutgers or leap or teacch or denver)	10
CMA practice guideline (infobase)		Autism	0
US National guideline clearinghouse		Autism	2 non-relevant
ECRI		Autism	1
Globe and Mail		(N/A, general scan)	1

- Date Limits: 1985 – 2000 when available
- Publication type limit to: NO limit
- Age Limit: no (to include theoretical and discussion materials)

The abstracts were reviewed. Only critical reviews were chosen. The term 'critical' refers to reviews that appraised the scientific validity of the primary research studies according to a set of criteria. Two critical reviews were found; one was by ECRI [15], the other was by BCOHTA [6]. In addition a few articles were received from Alberta Children's Services and one was a critical review of early intervention in autism by Smith [52]. No additional critical reviews were discovered. From the critical reviews, the original study articles were identified, but by the end of September not all were retrieved.

APPENDIX B: TREATMENT APPROACHES FOR CHILDREN WITH AUTISM

Heflin and Simpson [24] summarized various treatment approaches for children with ASD. These include interventions based on the formation of interpersonal relationships, behavioural or skills-based treatment programs, physiologically oriented intervention programs, and combined programs.

Interventions based on formation of interpersonal relationships:

Interventions based on the formation of interpersonal relationships include: "Holding Therapy", "Gentle Teaching", "Options", and "Floor Time". These therapies grew out of traditional psychoanalytic theory. The foundation for these therapies is the general belief that autism is due, in part, to an absence of warmth and caring from parents. Therapy should therefore involve traditional psychoanalytic interventions directed at assisting the autistic child to "uncover unconscious motives and repressed childhood conflicts to reconstruct his or her basic personality" [24]. Relationship-based approaches seek to assist the child to develop affect, bonding, and a sense of relatedness. Few studies were conducted to examine the empirical effectiveness of these therapies and the beliefs that formed the basis of these interventions are generally not accepted today.

Behavioural/skills-based treatment programs:

The intent of these programs is to develop or support the demonstration of specific skills rather than promote relatedness and attachment. Examples of these programs include:

- "Picture Exchange Communication System"(PECS) – this program uses pictures and other symbols to develop functional communication. No research studies are presented, but Heflin and Simpson [24] stated they believe that this program is an empirically sound method for developing communication skills.

- "Applied Behavior Analysis" (ABA) – Heflin and Simpson [24] noted that ABA grew out of behavior modification work. Individual analyses of a child's functioning are undertaken to identify and analyze skills needed for improved performance and functioning. Systematic teaching and intervention methods are used to train students to independently respond. Heflin and Simpson claim that ABA is "one of the most efficacious of intervention strategies for children with autism", however, it is also among the most controversial. The controversy revolves around outcome claims, exclusivity, extensive use and personnel. Issues of outcome are associated with the claim that ABA can lead to complete recovery. Issues around exclusivity pertain to whether ABA should be

used to the exclusion of other methods. Further issues are around the claim that 40 hours per week are necessary for several years for young children with autism. Finally, controversy exists around the use of "noneducationally" certified personnel to implement ABA. Lovaas therapy is provided as an example of ABA by the authors.

- "Cognitive Behavioural Methods" – a variety of behavioural and cognitive strategies have been used for higher functioning autistic children. Research on the efficacy of specific elements of these methods is lacking.

Physiologically oriented intervention program:

In contrast to the above programs, Heflin and Simpson [24] stated physiologically based programs are aimed at interventions that address neurological dysfunction believed to be involved in autistic disorders. These programs include:

- "Sensory Integration" (SI) – this program is aimed at assisting the child to organize and process sensory information. Proponents of SI include occupational therapists who help children with autism to process and use sensory information. Research on the efficacy of this approach is underway, however; thus far no studies have shown its effectiveness according to Heflin and Simpson.

- "Auditory Integration Training" (AIT) – this method of intervention seeks to reduce sound sensitivity and thereby improve behavioural, social and cognitive functioning. The general premise of this program is that children with autism have a sensory dysfunction, most likely a hypersensitivity to sound sensations, making a number of sounds painful to hear. AIT is not well accepted by the professional community. Heflin and Simpson cite only one research study and it does not provide support for the program.

- Psychopharmacologic Treatments – although not a cure for autism, many children with autism respond positively to psychotropic medications. These drugs must be used in combination with other treatments.

- Dietary Treatments – since the 1970's there have been various claims that megavitamin therapy is necessary for children with autism. No solid evidence has been documented to support this claim.

Combined programs:

These programs use combinations of the relationship-based, behavioural/skill-based and physiologically based approaches. One example is Treatment and Education of Autistic and related Communication handicapped Children (TEACCH). This program was started as a parent and child psychoanalysis

program. The original relationship-based approach changed to a behavioural/skill-based approach. The critical elements in TEACCH include "early identification, parent training, education, social and leisure skill development, communication training, and vocational preparation". The Childhood Autism Rating Scale (CARS) [47] and the Psychoeducational Profile – Revised (PEP-R) [46] to diagnose and gauge severity of autism were developed by the professional providers of the TEACCH program. The authors noted that the TEACCH program has been validated by a number of studies that "demonstrate the effectiveness of individual components of structured teaching".

APPENDIX C: OUTCOME MEASURES IN STUDIES REVIEWED BY ECRI, BCOHTA AND SMITH

Table 6 provides a list of outcome measures used in the studies reviewed by ECRI, BCOHTA and Smith. If possible, the primary study was requested to gather this information. Not all studies were available and the outcome measures listed were those noted by Smith [52].

Table 6: Outcome measures used in the primary studies

Study	Outcome Measure(s)
Anderson et al. [4] 1987	▪ Bayley Scales of Infant Development, or the Stanford-Binet Intelligence Scale for Children ▪ Vineland Social Maturity Scale or the Vineland Adaptive Behavior Scales ▪ A variety of instruments to obtain a measure of language performance; Symbolic Play Test, Peabody Picture Vocabulary Test, Pre-school Language Scale, Sequenced Inventory of Communication Development ▪ Uniform Performance Assessment System ▪ Parent satisfaction survey. ▪ Behavior Observations ▪ School placement ratings based on the integration with nonhandicapped children.
Birnbrauer and Leach [9] 1993	▪ Bayley Scales of Infant Development, Stanford-Binet Intelligence Scale, Leiter International Performance Scale, Peabody Picture Vocabulary Test, Wechsler Intelligence Scale for Children- Revised, Wechsler Pre-school & Primary Scale of Intelligence ▪ Vineland Adaptive Behavior Scale ▪ Reynell Developmental Language Scales, Receptive-Expressive Emergent Language Scale ▪ Personality Inventory for Children ▪ Parenting Stress Index ▪ Behavioural Observations
Fenske, et al. [16] 1985 *Outcome Measures as noted by Smith [52]	▪ Entry into public School
Handleman et al. [21] 1991 *Outcome Measures as noted by Smith [52]	▪ Battelle Developmental Inventory ▪ Learning Accomplishment Profile
Harris et al. [23] 1991 *Outcome Measures as noted by Smith [52]	▪ Stanford-Binet IV ▪ Pre-school Language Scales
Harris et al. [22] 1990 *Outcome Measures as noted by Smith [52]	▪ Pre-school Language Scales
Hoyson, Jamieson and Strain [27] 1984	▪ Learning Accomplishment Profile

Table 6: Outcome measures used in the primary studies (cont'd)

Study	Outcome Measure(s)
Jocelyn, et al. [28] 1998	• Leiter International Performance Scale • The Early Intervention Developmental Profile and the Pre-school Developmental Profile • Childhood Autism Rating Scale • Autism Behavior Checklist • Stress Arousal Checklist • TRE-Add Autism Quiz to evaluate parent's knowledge • The Family Assessment Measure • Client Satisfaction Questionnaire
Lord and Schopler [32] 1989 *Outcome Measures as noted by Smith [52]	• Merrill-Palmer Scale of Mental Tests, or Leiter International Performance Scales, or Weschsler Intelligence Scales for Children-Revised
Lovaas [35] 1987 McEachin, Smith and Lovaas [36] 1993	• Pre-Testing: Bayley Scales of Infant Development, or the Cattell Infant Intelligence Scale, or the Stanford-Binet Intelligence Scale, or the Gesell Infant Development Scale, or Vineland Social Maturity Scale (choice dependent on developmental level). • Post-Testing: Wechsler Intelligence Scale for Children- Revised, or the Merrill-Palmer Pre-school Performance Test, or the Peabody Picture Vocabulary Test or the Leiter International Performance Scale (tests choice depended on developmental level). • Vineland Adaptive Behavior Scales. • Personality Inventory for Children. • School Placement • Behavioural Observations of self-stimulatory behaviours, appropriate play behaviours, & recognizable words. • 1 hour parent interview
Ozonoff & Cathcart [37] 1998	• Psychoeducational Profile – Revised • Childhood Autism Rating Scale
Panerai, Ferrante and Caputo [38] 1997 *Not mentioned by ECRI	
Rogers and DiLalla [40] 1991	• Merrill-Palmer or Bayley Scales of Infant Development • Early Intervention Developmental Profile and Pre-school Profile • Childhood Autism Rating Scale • Observations of Play
Rogers, Lewis and Reis [43] 1989	• The Bayley Scales of Infant Development, or the Leiter International Performance Scale, or the Merrill Palmer Test of Mental Abilities (choice dependent on developmental level and age). • Early Intervention Profile and Pre-school Profile • Childhood Autism Rating Scale • Play Observation Scale
Rogers and Lewis [42] 1987	• Early Intervention Profile and Play school Profile. • Play school Observation Scale
Rogers et al. [41] 1986 *Outcome Measures as noted by Smith [52]	• Early Intervention Developmental Profile • Observations of Play

Table 6: Outcome measures used in the primary studies (cont'd)

Study	Outcome Measure(s)
Sheinkopf and Siegel [50] 1998 *Outcome Measures as noted by Smith [52]	▪ Merrill-Palmer Scale of Mental Tests
Short [51] 1984	▪ Merrill-Palmer Scale of Mental Tests, or Bayley Scales of Infant Development or the Wechsler Pre-school & Primary Scale of Intelligence ▪ Vineland Adaptive Behavior Scale ▪ Behavioural Observation of parental and child behavior ▪ Interviews with parents ▪ 2 questionnaires on effects of child problems on family, one to clinicians/ one to mothers.
Weiss [57] 1999 *Not mentioned by ECRI	

APPENDIX D: DIAGNOSTIC AND ASSESSMENT INSTRUMENTS APPROPRIATE FOR USE WITH CHILDREN WITH AUTISM/RESEARCH

Instrument	Area Assessed	Reliability/Validity	Comments
*Achenbach Child Behavior Checklist (ACBC) [2]	- completed by an adult it measures **externalizing** and internalizing behaviours.	- a well known and widely used measure of child behavior. - good reliability/validity.	- completed by the parents and teachers. - it has been used as a follow-up measure.
*Autism Diagnostic Interview – Revised (ADI-R) [31, 34]	- semi-structured interview for caregivers of autistic children. - used for **diagnostic purposes.** - considered by some as a measure of high diagnostic accuracy. - one of the better standardized instruments for establishing the diagnosis of autism.	- reliability was tested and "indicate that the ADI-R is a reliable and valid instrument" - no ratings provided	- the ADI-R is considered by some professionals as a measure of high diagnostic accuracy. - assessment begins with a home visit.
Autism Screening Instrument for Educational Planning [29] (ASIEP-2)	- provides data on 5 aspects of behavior – sensory, relating, body-concept, language and social self-help. - provides a profile of abilities in verbal behavior, social interaction, educational level and learning characteristics. - used primarily for **evaluating & planning.**	- only reported as a "strong intercorrelations among the ASIEP-2 subtests and the utility of the battery to distinguish among groups of subjects with a variety of disabilities"	- one of the most popular individual assessment instruments available. It has been researched around the world. - standardized & researched in diagnostic centers.
Autism Treatment of Evaluation Checklist *Copies available @ www.autism.com/atec	- measures the effectiveness of various treatments. - stated to measure changes in response to treatment. - primary function is to measure the **efficacy of interventions.**	- reliability = 0.815 - 0.920 for subscales, 0.942 for total score	- one page designed to be completed by parents, teachers & caretakers. Consists of 4 subtests – Speech/Language Communication, Sociability, Sensory/Cognitive Awareness & Health/Physical Behavior - not copyrighted & may be used free by researchers

Intensive intervention programs for children with autism

Appendix D: Diagnostic and assessment instruments appropriate for use with children with autism/research (cont'd)

Instrument	Area Assessed	Reliability/Validity	Comments
*Bayley Scales of Infant Development [7]	- designed to identify children who have a **cognitive or motor delay** and suggests needed forms of intervention. - evaluates a large variety of abilities – sensory perceptual acuities, discriminations, memory, learning, problem solving, mental mapping, complex communication, vocalization, verbal language & mathematical concept formation. There is also a motor scale.	- renormed recently on a large US sample. - this is a standardized test, widely used in clinical and research.	- recently revised. - yields a normalized standard score called the Mental Developmental Index
*Childhood Autism Rating Scale (CARS) [14, 47, 48] *Available from Western Psychological Services	- developed to formalize observations of autistic children's behavior throughout the day. - used primarily for **diagnostic purposes.** - brief, convenient, and suitable for use with any child >2years.	- not reported.	- children are rated on a 7 point scale to determine how much behavior deviates from "normal". - can be used by any professionals who are trained to use it.
*Columbia Mental Maturity Scale (CMMS-III) [11, 44]	- measures **nonverbal intelligence** & is useful to evaluate children who have sensory or motor defects, or who have difficulty speaking, and to some extent reading. - the scale provides a means for evaluating children with disabilities and may be less culturally loaded than some other intelligence tests	- is a frequently used test that is reliable and valid for this population.	- scores are not interchangeable with those obtained by the SBIS-IV, WICS-III or WPPSI-R. - it appears to measure general reasoning ability.
*Expressive One-Word Picture Vocabulary Test and Receptive One-Word Picture Vocabulary Test [18]	- measures the child's ability to **verbally label objects** and people. - measures one word-hearing vocabulary and provides a measure of the child's ability to understand language.	- standardized	- this is a standardized test that provides age equivalents, standard scores, scaled scores, percentile ranks and stanines.

Intensive intervention programs for children with autism

Appendix D: Diagnostic and assessment instruments appropriate for use with children with autism/research (cont'd)

Instrument	Area Assessed	Reliability/Validity	Comments
Gilliam Autism Rating Scale (GARS) [19]	- helps identify and **diagnose** autism. Based on the DSM-IV. - offers a developmental approach to the assessment of children with autism. - it is an inventory of behaviours and skills designed to identify uneven and idiosyncratic learning patterns.	- the GARS is reported to have strong reliability and validity. The test was normed with 1,092 subjects with autism.	- can be completed by teachers, parent and professionals. - it takes 45 minutes to 1.5 hours to administer.
*Leiter International Performance Scale (LIPS) [30]	- measures **intelligence independent of language ability** for children 3 and older. - can be used as an aid in clinical diagnosis, especially if testing children who can not speak	- norms are outdated, standardization is inadequate, lacks information about reliability of the scale for various age groups.	- tends to underestimate intelligence.
*Merrill Palmer Scale of Mental Tests (MPSMT) [26]	- widely used nonverbal test for assessing **visual -spatial skills.** - can be used for autistic children as this is usually an area of strength for these children.	- is a frequently used test with high reliability and validity.	- the MPSMT enables a more detailed assessment of visual-perceptual functioning than is provided by the BSID-II or WPPSI-R.
*Parenting Stress Inventory (PSI-III) [7, 8]	- identifies **stressful areas in parent child interactions.** It is administered to the parents. It is a screening and diagnostic instrument that assumes that the total stress a parent experiences is a function of child characteristics, parent characteristics and situations of parenting.	- it is a widely used test with good reliability/ validity.	- there is a short form that takes 10 minutes to complete. - it is helpful in assessing early identification of dysfunctional parent-child relationships.
*Peabody Picture Vocabulary Test (PPVT-R) [13] *Available from the American Guidance Service, Circle Pines, Minn.	- measures receptive **vocabulary – one facet of general intelligence** - may be used as a screening device.	- widely used, standardized test with good reliability & validity.	- short time to administer, may be used as a screening tool

Intensive intervention programs for children with autism

Appendix D: Diagnostic and assessment instruments appropriate for use with children with autism/research (cont'd)

Instrument	Area Assessed	Reliability/Validity	Comments
Psychoeducational Profile – Revised (PEP-R) [47]	- offers a developmental approach to the assessment of children with autism. It is an inventory of behaviours and skills designed to identify uneven & idiosyncratic learning patterns. - provides information on **developmental functioning** in imitation, perception, fine motor, gross motor, eye-hand integration, cognitive performance and cognitive verbal areas.	- not reported	- the PEP-R kit consists of a set of toys and learning material presented with structured play activities. - total time to administer = 45 min to 1.5 hrs.
Questionnaire on Resource & Stress (QRS) [25, 49]	- this questionnaire measures **stress in parenting.**	- the questionnaire fulfills requirements for an acceptable level of validity.	- used widely in research studies. - there is a short form available.
Real Life Rating Scale [17]	- used to assess the effects of treatment on 47 variables – motor, social, affective, language and sensory domains. - can be administered by non-professionals, is rapidly scored by hand, and can be repeated frequently.	- not reported	- can be repeated frequently without affecting inter-observer agreement. - can be done in natural settings by non-professional raters. - rapidly scored by hand.
Reynell Developmental Language Scales [39]	- measures **comprehension and expressive language**	- is well known as a standardized, reliable test.	- one of the most common tests for language delayed children.
Sequenced Inventory of Communication Development (SICD-R) *Available from University of Washington Press, Seattle	- tests a variety of **early communication skills** giving a broad measure of the semantic, syntactic, and pragmatic aspects of a child's receptive and expressive language skills.	- no information given	- the test combines behavioural items with parental report items.

Intensive intervention programs for children with autism

Appendix D: Diagnostic and assessment instruments appropriate for use with children with autism/research (cont'd)

Instrument	Area Assessed	Reliability/Validity	Comments
*Stanford-Binet Intelligence Scale (SBIS-IV) [54]	- **intelligence test** used for ages 2 to adult.	- recently revised and a new national standardization completed. - is a widely used clinical and research instrument that is very reliable. - is regarded by some as a gold standard	- used clinically and in research. - takes 30 minutes to 4 hours as all tests do not have to be administered.
*Test of Nonverbal Intelligence (TONI-II) [10]	- is a language free measure of **cognitive ability**. It measures abstract figural problem solving in children 5 and older. - it is an ideal test for assessing individuals with speech, language or hearing impairments, those who have suffered brain damage, or have academic handicaps, and those who do not speak English	- a standardized measure that has good reliability and validity for normal, mentally retarded, learning disabled, deaf and gifted children.	- two equivalent forms make the TONI-II ideal for situations where pre and post measures are desirable.
*Vineland Adaptive Behavior Scales (VABS) [53]	- measures overall adaptive behavior in areas of communication, daily living skills, socialization, motor skills & maladaptive behaviours. - it is administered by interviewing the child's parents, teachers, or caregivers.	- questions have been raised about the accuracy of standard scores across the age range, there is a lack of uniformity of scores across various ages.	- one of the most frequently reported measures in studies reviewed. - there is considerable overlap among the domains with both communication and daily living domains containing questions about the child's language ability.
*Wechsler Pre-school & Primary Scale of Intelligence (WPPSI-R) [55]	- **intelligence test** for assessment in pre-school children.	- is highly reliable, standardized, valid. Stated to be the gold standard.	- subtest and total scale scores derived. - has norms for 17 age groups divided by three month intervals. -testing takes 1-4 hours.
*Wechsler Intelligence Scale for Children (WISC-III) [56]	- **intelligence test** for school age children - it is a valuable tool for psychoeducational assessment, diagnosis, placement and planning	- is highly reliable, standardized, valid. Stated to be the gold standard.	- it can be used to diagnose and has a strong place in neuropsychological assessment and in research. -testing takes 1-4 hours.

* Test use or interpretation is restricted to those who have knowledge of the theory and methodology of standardized assessment procedures and supervised training in working with parents and children.

REFERENCES

1. Abidin R. *Parenting stress index.* Charlottesville: Pediatric Psychology Press. 1983.

2. Achenbach TM. *Intregrative Guide to the 1991 CBCL TSR and TRF profiles.* University of Vermont, Burlington. 1991.

3. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders. DSM-IV(4th ed.).* Washington, D.C.: American Psychiatric Association, 1994.

4. Anderson SR, Avery DL, DiPietro EK, et al. Intensive home-based early intervention with autistic children. *Education & Treatment of Children* 1987;10(4):352-66.

5. Autism Society of America. Intensive Behavioural Intervention. *http://www.autism-society.org/packages/discrete_trial.html.* 2000.

6. Bassett K, Green CJ, Kazanjian A. *Autism and Lovaas treatment: a systematic review of effectiveness evidence.* Vancouver, B.C.: British Columbia Office for Health Technology Assessment Report 00:IT. 2000.

7. Bayley N. *Manual for the Bayley scales of infant development.* New York: The Psycho-logical Corporation. 1969.

8. Berry J, Jones W. The parental stress scale: Initial psychometric evidence. *Journal of Social & Personal Relationships* 1995;12(3):463-72.

9. Birnbauer JS, Leach DJ. The Murdoch early intervention program after 2 years. *Behavior Change* 1993;10:63-74.

10. Brown L, Sherbenoa RJ, Johnsen SK. *Test of nonverbal intelligence.* Austin, TX: Pro-Ed. 1990.

11. Burgemenster B, Blum L, Lorge I. *Columbia Mental Maturity Scale (3rd. ed.).* New York: Harcourt Brace. 1972.

12. California Departments of Education and Developmental Services. Collaborative Work Group on Autistic Spectrum Disorders. Best practices for designing and delivering effective programs for individuals with autistic spectrum disorders. *http://www.bunt.com* 1997.

13. Dunn L, Dunn LM. *Peabody picture vocabulary test-revised: Manual for forms L and M.* Circle Pine, MN: American Guidance Services. 1981.

14. Eaves RC, Milner B. The criterion-related validity of the childhood autism rating scale and the autism behavior checklist. *Journal of Abnormal Child Psychology* 1993;21(5):481-91.

15. ECRI. *Comprehensive programs for the treatment of children with autism.* Plymouth Meeting, PA: ECRI Report, 2000.

16. Fenske E, Zalenski S, Krantz PJ, McClannahan LE. Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis & Intervention in Developmental Disabilities* 1985;5(1-2):49-58.

17. Freeman B, Rivto E, Yokota A, Rivto A. A scale for rating symptoms of patients with the syndrome of autism in real life testings. *Journal of the American Academy of Child Psychiatry* 1986;25(1):130-36.

18. Gardner M. *Expressive one-word picture vocabulary test.* Austin, TX: Pro-Ed. 1990.

19. Gilliam J. *Gilliam autism rating scale.* Austin, TX: Pro-Ed. 1995.

20. Gresham FM, MacMillan DL Autistic recovery? An analysis and critique of the empirical evidence on the Early Intervention Project. *Behavioural Disorders* 1997;22(4):185-201.

21. Handelman JS, Harris SL, Celiberti D, et al. Developmental changes of pre-school children with autism and normally developing peers. *Infant-Toddler Intervention: the Transdisciplinary Journal* 1991;1:137-43.

22. Harris S, Handleman J, Gordon R, Kristoff B. Changes in cognitive and language functioning of pre-school children with autism. *Journal of Autism & Developmental Disorders* 1991;21(3):281-90.

23. Harris S, Handleman JS, Kristoff B, et al. Changes in language development among autistic and peer children in segregated and integrated pre-school settings. *Journal of Autism & Developmental Disorders* 1990;20(1):23-31.

24. Heflin LJ, Simpson RL. Interventions for children and youth with autism: prudent choices in a world of exaggerated claims and empty promises. Part I: Intervention and treatment options review. *Focus on Autism and Other Developmental Disabilities* 1998;13(4):194-211.

25. Holyrod J. *Questionnaire on resources and stress.* Brandon, VT: Clinical Psychology Publishing Company, Inc. 1987.

26. Howlin PA, Rutter M. *Treatment of autistic children.* New York: Wiley Publishers. 1987.

27. Hoyson M, Jamieson B, Strain PS. Individualized group instruction of normally developing and autistic-like children. The LEAP communication model. *Journal of the Division for Early Childhood* 1984;157-72.

28. Jocelyn LJ, Casiro OG, Beattie D, et al. Treatment of children with autism: a randomized controlled trial to evaluate a caregiver-based intervention program in community day-care centers. *Journal of Developmental Behaviour Pediatric* 1998;19(5):326-34.

29. Krug D, Arick J, Almond P. *Autism screening instrument for educational planning (2nd. ed.).* Austin, TX: Pro-Ed. 1993.

30. Leiter R. Part I of the manual for the 1948 revision of the Leiter international performance scale: Evidence of the reliability and validity of the Leiter tests. *Psychological Service Center Journal* 1959;11:1-72.

31. Lord C, Rutter M, LeCouteur A. Autism diagnostic interview - revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism & Developmental Disorders* 1994;24(5):659-85.

32. Lord C, Schopler E. The role of age at assessment, developmental level, and test in the stability of intelligence scores in young autistic children. *Journal of Autism & Developmental Disorders* 1989;19(4):483-99.

33. Lord C, Schopler E. TEACCH services for pre-school children. In: *Pre-school education programs for children with autism.* Austin, TX: Pro-Ed., 1994:87-106.

34. Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in pre-school children. *Infant Mental Health Journal* 1993;14(3):234-52.

35. Lovaas OI. Behavioural treatment and normal educational and intellectual functioning in young autistic children. *Consulting and Clinical Psychology* 1987;55(1):3-9

36. McEachin JJ, Smith T, Lovaas OI. Long-term outcome for children with autism who received early intensive behavioural treatment. *American Journal of Mental Retardation* 1993;97(4):359-72.

37. Ozonoff S, Cathcart K. Effectiveness of a home program intervention for young children with autism. *Journal of Autism & Developmental Disorders* 1998;28(1):25-32.

38. Panerai S, Ferrante L, Caputo V. The TEACCH strategy in mentally retarded children with autism: a multidimensional assessment: pilot study. *Journal of Autism & Developmental Disorders* 1997;27(3):345-47.

39. Reynell J, Gruber C. *Reynell developmental language scales.* Los Angeles, CA: Western Psychological Association. 1990.

40. Rogers SJ, DiLalla D. A comparative study of a developmentally based pre-school curriculum on young children with autism and young children with other disorders of behavior and development. *Topics in Early Child Special Education*; 1991;11(2):29-47.

41. Rogers SJ, Herbison J, Lewis H, et al. An approach for enhancing the symbolic, communicative, and interpersonal functioning of young children with autism or severe emotional handicaps. *Journal of the Division for Early Childhood* 1986;10(2):135-45.

42. Rogers SJ, Lewis H. An effective day treatment model for young children with pervasive developmental disorders. *Journal American Academy Child & Adolescent Psychiatry* 1989;28(2):207-14.

43. Rogers SJ, Lewis H, Reis K. An effective procedure for training early special education teams to implement a model program. *Journal of the Division for Early Childhood* 1987;11(2):180-88.

44. Rueter J, Mintz J. Columbia mental maturity scale as a test of concept formation. *Journal of Consulting & Clinical Psychology* 1970;34:387-93.

45. Sattler J. *Assessment of children: Revised and updated third edition.* San Diego, CA: Jerome M. Sattler Publishers Inc. 1992.

46. Schopler E, Reichler R, Bashford A, et al. *Psycholeducational profile (rev.).* Austin, TX: Pro-Ed. 1990.

47. Schopler E, Reichler RJ, Renner B. *The childhood autism rating scale (CARS): for diagnostic screening and classification of autism.* New York: Irvington Publishers 1986.

48. Schopler E, Reichler RJ, Renner B. CARS (*The childhood autism rating scale*). Los Angeles: Western Psychological Services 1988.

49. Scott R, Sexton D, Thompson B, Wood T. Measurement characteristics of a short form of the questionnaire on resources and stress. *American Journal of Mental Retardation* 1989;94(3):331-39.

50. Sheinkopf SJ, Siegel B. Home-based behavioural treatment of young children with autism. *Journal of Autism & Developmental Disorders* 1998;28(1):15-23.

51. Short A. Short-term treatment outcome using parents as co-therapists for their own autistic children. *Journal of Child Psychology & Psychiatry* 1984;25(3):443-58.

52. Smith T. Outcome of early intervention for children with autism. *Clinical Psychology - Science & Practice* 1999;6(1):33-49.

53. Sparrow S, Balla D, Cicchetti D. *Vineland adaptive behavior scale (4th ed.).* Circle Pines, MN: American Guidance Services 1984.

54. Thorndike R, Hagen E, Sattler M. *The Stanford-Binet intelligence scale (4th ed.).* Chicago, ILL: The Riverside Publishing Company 1986.

55. Wechsler D. *Manual for the Wechsler pre-school and primary scale of intelligence (rev).* San Antonio, TX: The Psychological Corporation. 1989.

56. Wechsler D. *Manual for the Wechsler intelligence scale for children (3rd. ed.).* San Antonio, TX: The Psychological Corporation. 1991.

57. Weiss MJ. Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Intervention* 1999;14(1):3-22.